

Data modeling for museum collections

Original

Data modeling for museum collections / Lo Turco, Massimiliano; Calvano, Michele; Giovannini, Elisabetta Caterina. - In: INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES. - ISSN 2194-9034. - ELETTRONICO. - XLII-2/W9:(2019), pp. 433-440. [10.5194/isprs-archives-XLII-2-W9-433-2019]

Availability:

This version is available at: 11583/2724891 since: 2020-07-10T10:56:17Z

Publisher:

Copernicus Publ.

Published

DOI:10.5194/isprs-archives-XLII-2-W9-433-2019

Terms of use:

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

DATA MODELING FOR MUSEUM COLLECTIONS

M. Lo Turco *, M. Calvano *, E. C. Giovannini *

* DAD, Department of Architecture and Design, Politecnico di Torino, Italy
(massimiliano.loturco, michele.calvano, elisabetta.caterina.giovannini)@polito.it

Commission II

KEY WORDS: 3D Modelling, Data Modeling, Museum Collections, Archival Research, Paradata Documentation

ABSTRACT:

The relationship between cultural heritage, digital technologies and visual models involves an increasingly wide area of research, oriented towards the renewal of archives and museums for the preservation and promotion of culture. Recent research activities are the result of the progressive strengthening of digital technologies and the needs of a new generation of "digital" users, which requires museums to update their means of communication using Semantic Web languages and technologies shaped by a social conceptualization of a graph-based representation of information.

The growth of several digitized heritage collections increases the necessity of proper methodologies to develop a structured system able to access to these collections and the large amount of data, metadata and paradata related to the digitized objects in a structured and organized way, defining a set of collection information models (CIM), that considers not only the digitizing process but also the data collection process, layered by an Upper Ontology level structure, based on CIDOC-CRM.

1. INTRODUCTION

The paper introduces a possible workflow with the purpose to define a methodology that could take into account new possibilities on publishing museum "hidden" collections on the web, facing with the necessity to create, annotate, assert, argue, search, cite, and justify data-driven research.

To create digital contents and publish them on the web, a series of considerations about description and visualization of data, is necessary. The museum community already developed a series

of data-model to describe information and share knowledge about digital collections, but a common procedure is still missing.

The development of a data-model that can meet the needs of a museum digital collections should be able to fix the following needs:

- (a) cataloguing (collection, acquisition and conservation management);
- (b) publication of metadata and paradata (presentation and visualization of data for dissemination purpose);
- (c) portals and system management for museum operators.

Then, this project focused the attention on the possibility to visualize a set of collection information models (CIM) considering museum necessity, but with a multidisciplinary point of view: develop a workflow that could be useful for historians and 3d modelers, open to academics research.

This approach, in fact, considers not only the digitizing processes, but also the data collection and modelling phases.

The data structure has been developed following CIDOC-CRM (last released version 6.2.4) (Le Boeuf, Doerr, Emil Ore, & Stead, 2018) as a formal and generic structure of concepts and relationships with the aim to propose an Upper-level Ontology to set up information and address the complexity of museum collection object data management.

1.1 Background

In the last decade the web scenario evolved from the web 2.0 phase into a 3.0 era, which enabled users to participate in the creation, sharing and aggregation of web contents (O'reilly, 2007). Even if Markoff was the first to propose a definition of the "web 3.0" (Markoff, 2006) several authors identified this phase with the Semantic Web that could be considered one of its primary components.

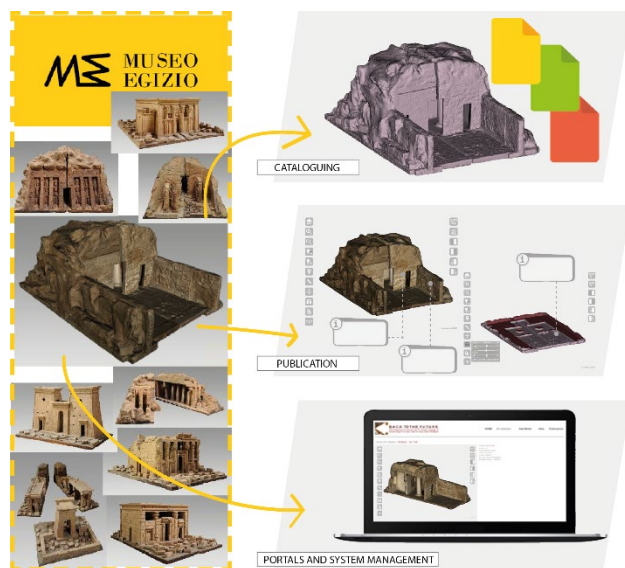


Figure 1. data-model structure for museum digital collection needs

* Although the contribution was conceived jointly, Massimiliano lo Turco is author of paragraphs 1 and 4; Michele Calvano of paragraph 3; Elisabetta C. Giovannini of paragraphs 1.1, 1.2 and 2.

After the diffusion of the Web 2.0, many applications were developed, and with them many data formats and query languages focusing the attention on semantics of information. Technological key elements that are nowadays recognizable in the Web are:

- a) eXtensible Markup Language (XML) a standard format for data exchange;
- b) Web Application Programming Interfaces (APIs) that are methods to request data from a web server in order to use them in external applications;
- c) Web services standard protocols and architectures specified by the World Wide Web Community (W3C);

In this scenario the idea of a "Web of Data" (Berners-Lee, 2006) follows the idea of linking contents (data) embedded in documents (HTML pages) in a global information space: the Semantic Web, that "is not a separate web but an extension of the current one, in which informations is given well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee, Hendler, & Lassila, 2001)

In the "Semantic Web" the need to express data and information on the Web introduce the Uniform Resource Identifier (URI) that identify the representation of a resource and a new W3C model: Resource Description Framework (RDF), an XML based language that deal with semantic interoperability and that it is made by three components univocally identified by an URI. Finally, ontologies represents the most advanced way of representing knowledge on the web, enriching RDF Schema (RDFS), defining relationships between concepts. (Bruseker, Carboni, & Anais, 2017)

1.1.1 The semantic web and Cultural Heritage

A serious number of instruments to model informations and manage cultural heritage data were developed. Many initiatives tried to create systems able to describe different kind of resources in an electronic environment trying to apply benefits of Semantic Web technologies to Cultural Heritage: Dublin Core (DC) metadata Elements and DC Terms (Powell, Nilsson, Naev, Johnston, & Baker, 2007), Simple Knowledge Organization System (SKOS) (Miles & Bechhofer, 2009), Functional Requirements for Bibliographic Record (FRBR) (Tillett, 2005), Europeana Data Model (EDM) (Meghini et al., 2016), MIDAS Heritage standard (Forum on Information Standards in Heritage (FISH), 2012), Lightweight Information Describing Objects (LIDO) (Coburn, Light, McKenna, Stein, & Vitzthum, 2010), VRA Core (Masci, 2009).

In the case of digital collections, data modeling challenges has to face with a huge amount of data available in multiple formats (images, texts, 3d models, etc.) that are not only multi-targeted (museum operators, tourists, academics, etc.) but also multi-disciplinary (archaeology, history, architecture, science, etc.).

The growing of digitization and sharing of museum collections makes necessary to organize databases based on an appropriate computer ontology, i.e. a series of conceptual distinctions - that

are impossible without the development of a semantics and a vocabulary provided to computers by (human) programmers - that are transversal and interoperable.

It is commonly recognized that the diversity of museum collections databases is the main cause of problems on information retrieval. This is the reason why in 1999 the International Committee for Documentation (CIDOC) developed a Conceptual Reference Model (CRM) (Le Boeuf et al., 2018) to solve engineering problem of knowledge integration across museum databases, faced by the International Council of Museums (ICOM).

1.2 The digital collection Nubian Temple's maquettes

The examined case study concerns some wooden models (maquettes) that are in-scale reproductions of the Nubian temples of the ancient Egypt, mainly preserved in the depots of the Egyptian Museum in Turin.

At the regard, the data collection can cover different dimensions of object knowledge complexity that can be summarized as follows:

1) Documentation of the object:

a) Historical-artistic data about the wooden model as objects that belong to a museum collection: the information related to the scale of reproduction and the value possessed by the artefact in relation to the historical events that it documents were taken into account Data for the maintenance and the management of the object in relation to the places where it is exhibited: these information are generally included in the collection management systems of museums.

b) Digital acquisition of the object: the result is a multifaceted representation (mesh) with an optimized topology for web diffusion (3D modeling).

2) Historical-artistic documentation related to the architecture represented by wooden maquette, from which it is possible to access to the graphic, textual and numerical information that refers to the history of the monuments (historical images, views of travelers, previous investigations, procedures of the UNESCO campaigns for the recovery of temples, new places, etc....).

The digitization of both the artefact as the data archive related to it, allows to enrich the virtual models with technical and specialized information, as a possible way to rethink the usual procedures of filling in the forms, making the process of cataloguing and dissemination more efficient.

In the first phase of mapping, the archive data available were mapped using CIDOC-CRM classes and properties. Finally the CRMdig extension of CIDOC was used for the mapping of documentation of the evaluative, analytical, deductive, interpretative and creative decisions first made in the course of acquisition and then related to computer-based visualization. The CRMdig is an ontology realized to encode metadata and paradata about the steps and methods of production ("provenance") of digitization products and digital representations created by various technologies. (Doerr, Stead, & Theodoridou, 2015)

2. DATA MODELING

According with documentation about previous projects (Mantegari, 2009), the stages characterizing the CIDOC-CRM mapping workflow can be summarized as follows:

(1) Selection of relevant data sources with reference to the scope and goals of the case study (2) Analysis of the data schema (3) eventual selection of relevant subset of the data structure according to the identified events and processes (4) identification of the events and processes that are implicitly or explicitly represented in the data schema (6) grouping of data (7) mapping of each db field to the CRM classes and properties (8) evaluation

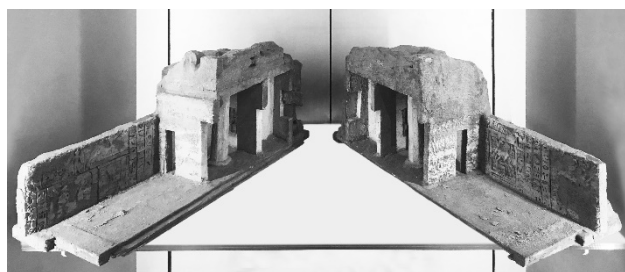


Figure 2. Beit el-Wali Temple Maquette



Figure 8. Above: Beit el Wali, digital relief of the model (dimensions of the two halves 51 x 68 x 29 cm). Below: Construction of the images by projection on the bounding box.



Figure 9. Elevation and cross-section of the surveyed model

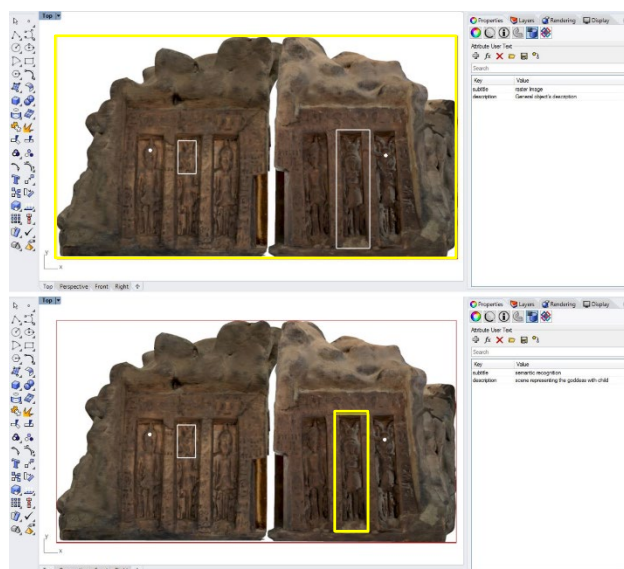


Figure 10. Annotation of geometry is done by following a hierarchy

to redraw more complex graphic conditions, giving space to the possible shapes.

In the visualization phase of the models and the relative information, open-source software is used for the creation of interactive 3D models that can be visualized in a Web environment (3DHOP in our experimentation).

The interoperability between the two environments is delegated to a visual programming language (Grasshopper in our experimentation), which is an open and easily implemented research tool.

In the work path expressed in this paper, the annotation phase is mainly delegated to expert users who, through simplified processes and in dedicated environments, relate disseminating data to pictures of the model and therefore to the model itself. Particular attention is given to the geometric objects used for the annotative process: this will be done in the plan, then in 2D, through the use of points and polylines that identify portions of the image. This is to simplify as much as possible the graphic operation for the construction of annotations, mainly shifting the attention of the expert user to the type and the quality of data.

3.2 Direct annotations

The first two operations, even though they are part of a sequential workflow, have the characteristic of being able to be carried out remotely, separated from a continuous but hierarchically determined workflow.

The possibility to untie the phase of images preparation (technical operation) and the annotation phase (for experts) allows a collaboration between heterogeneous users, linked by the common purpose of an increased knowledge of the acquired models. The working times are as follows:

- 1) Creating images from the model;
- 2) Annotations on images;
- 3) Remapping 2D annotations in 3D;
- 4) Implementazione del codice html di 3DHOP per le parti reiterative.

3.2.1 Creating images from the model

The annotative phase takes place in the plane, thus on two-dimensional images of the model. The images used are the result of the 3D model projection on the faces of its bounding box (Fig. 8). The process is automated thanks to procedures developed in VPL (Tedeschi & Andreani, 2014) for which, once the projections on the box surfaces are constructed, they are saved separately and then opened in any CAD software that has tools for annotation and data enrichment of the geometries. For complex models, characterized by internal spaces to be annotated, the automated procedure is able to produce section images (Fig. 9).

3.2.2 Annotations on images

The images are then reopened in the 2D space of a drawing software where there is the possibility to identify points and areas of different complexity, building boundaries by drawing parallelograms but also closed polylines. The freedom in the use of shapes allows a greater adherence to objects for identification or annotation.

The CAD software used is Rhinoceros 6, a modeling tool commonly used in the area of representation even in the academic field, as it is software that can easily be used for experimentation and research. The objective is to develop an annotation procedure that can be carried out on different drawing software whose minimum requirement is to have tools

for information enrichment and allows the manipulation of objects through programming languages.

In figure 10, the annotation operation is illustrated on a frontal image of the minor temple of Abu Simbel; on the basis of what is described at the beginning of this paragraph, we can identify an informative hierarchy in relation to the methods of annotation.

Through the selection of the objects present in the scene, it is possible to open an information enrichment space in which the expert user inserts the information he considers useful for a better understanding of the object. In the case in question, the information introduced is of a historical-artistic nature that frames the model surveyed in a historical moment and defines its cultural value.

The operation is organized in a hierarchical way: selecting the image opens a window of annotation in which to insert information related to the nature of the model: attribution, historical period of construction, materials, routes of the object, etc.. (Fig. 10 above).

For an annotation also of a semantic nature, a closed polyline is introduced, even if irregular, in order to frame portions of the image to which the information inserted in the information box will refer (Fig. 10 below). Annotation strings can also be filled in with links to multimedia content and images.

3.2.3 Remapping 2D annotations in 3D

The procedure proposes an annotation phase on two-dimensional images using a CAD equipped with tools for information enrichment, a phase that can take place offline and with proprietary tools and software.

The annotation phase is aimed at the compilation of an application (3DHOP in this search) for the navigation of 3D informed models in a web space. Therefore, the attributes previously associated with flat elements must be relocated on interactive geometries present in the model, which in the same way as the annotations, must have a semantic feedback without losing the information. Information that will be activated when the user selects, in the web portal, the interactive elements created.

The remapping of the geometries and information from the 2D image to the acquired 3D model takes place using operations that can be traced back to the Nurbs (Di Marco, 2017) mathematics of the scene built in the modeler and also by using the principles of descriptive geometry (Fig. 11)(Migliari, 2008) The images are orthogonal projections to the faces of the bounding box that circumscribes the 3D object. Let's think of the copies of the faces but oriented on the horizontal plane (XY

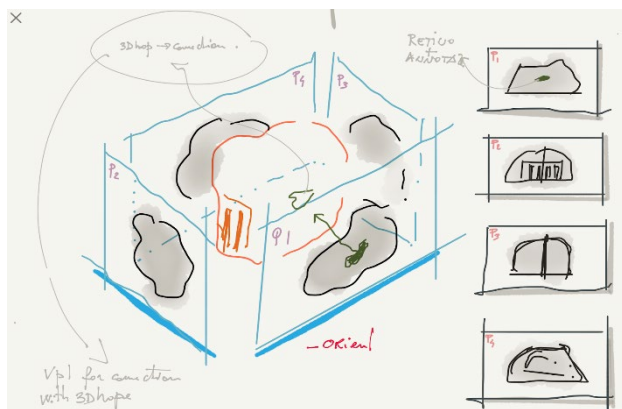


Figure 11. Remapping of the geometries and information from the 2D image to the acquired 3D model

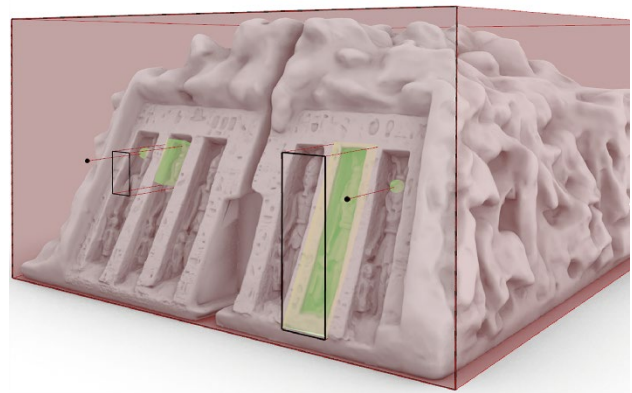


Figure 12. Remapped geometry from the picture, then projected on 3D model plane of the absolute reference system), but now with enriched

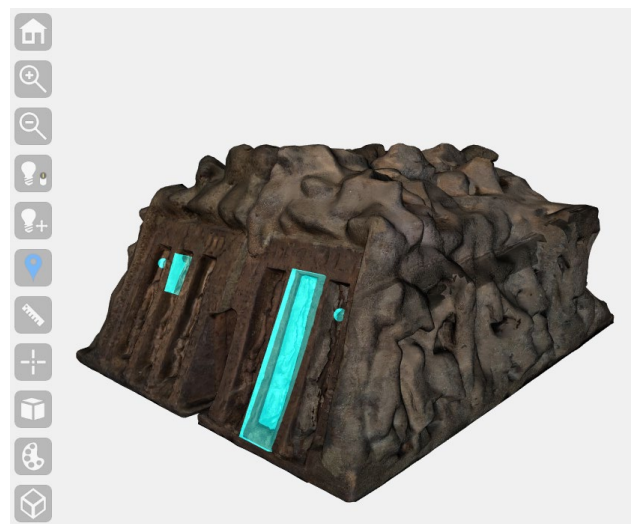


Figure 13. Hotspots layer on the web-portal interface

geometries provided by expert users as described in the previous paragraphs (points and closed polylines).

The images are intended as textured surfaces; the domain of surfaces oriented on the horizontal plane is similar to that of the faces of the bounding box from which the images were generated. Considering that the surfaces are a reference system $R2$ (two-dimensional) with two variables (u , v), then each geometric object on them is described by parametric coordinates u , v . By construction, the systems oriented in the plane (the images for the annotations) are identical or proportionate to the intrinsic systems of the faces of the bounding box; it is therefore possible to remap the coordinates of the geometric annotations bringing them back from the image to the faces of the bounding box.

The vertices of the semantic annotations (those generated by closed polylines) are interpolated by construction plans used as systems that orient new bounding box (called hotspots in the research); the latter contain the portions of the digital model affected by semantic recognition (Fig. 12). Future up-upgrades of the definition foresee the possibility to create point grids inside the annotative polygons that, projected on the model, will describe a structured point cloud to be interpolated with 3D primitives able to better discretize the highlighted model parts. Differently, the precise annotations projected on the model become centers of spheres (called spots in the research) associated with the information described in the previous phase.

In 3DHOP the models to be communicated through information enrichment are represented by multiresolution meshes. Therefore, the quality of the model increases in relation to the proximity of the camera that investigates the geometries. This is not the case for the spots and hotspots generated through the annotative process. In fact, 3D solids have to be simply exported in .ply and saved in the web server. Through appropriate IDs assigned to the generated models, the html code of the web portal

links the annotative information to the 3D elements. In this way, the model can be orbited and clicked on the spots and hotspots highlighting, in special pop-up windows, the information written in previous moments. In addition, as mentioned above, it is possible to attach links to images and multimedia files to annotative objects.

3.2.4 Implementation of 3DHOP html code for repeating parts

The code for the graphic construction of the web portal consists of a part for the implementation of the layout of the web portal (Fig. 13) and another part for the implementation of information related to annotative objects.

This last part can be defined as "parametric" because it is related to the number of annotations done on the images. As the number of annotations increases, the length of the code that manages the 3D objects of the portal varies, connecting them to the notes described by the expert.

For this reason, the basic code for the construction of the portal is initially discretized in the following parts:

- a) the part dedicated to the organization of the interface layout;
- b) the part describing the attributes and behaviors of the models in the scene.

The portions of code belonging to the second point, once mapped in relation to the basic syntax, can be replicated in relation to the number of annotative objects previously collected and associated in the CAD environment.

The replication of code parts is an action that brings technical difficulties, so a contextual goal of the research is to implement some procedures to facilitate the interoperability between the time of annotation and the time of display of models and information.

Through the use of the VPL language, a textual mask structured according to the html syntax has been created. The mask is composed of fragments of code to be repeated in relation to the number of annotative geometries introduced. Inside the code are placed special characters as placeholders to identify variables in which to place progressive elements to create the consequentiality of the code parts (Fig. 14).

At this time the code is proposed as an open VPL diagram allowing the usual changes and updates during the research phase; once validated the procedure on multiple case studies, the code will be clustered into individual components in order to become an add-on for VPL at the disposal of future operations of interrelation between CAD space and web portal.

The result of the automation of the annotative enrichment phase, allows the almost complete compilation of the part of the code delegated to describe the attributes and behaviors of the models in the scene and generated by the 2D annotations.

4. CONCLUSION AND FUTURE OUTCOMES

The illustrated activity is part of a wider project: first of all it aims to think about the digitization and semantic enrichment of some museum collections closed in the depots of the Egyptian Museum. In accordance with the recent BIM platforms used for the modeling of the building heritage, we worked on the definition of a workflow that would optimize the processes of information implementation of digital models. The outcome of these working phase were named as CIM models (Collection Information Modeling).

In analogy to the building field, where it is necessary that the different actors of the building process can share and exchange data on the different disciplines involved, also in this research it was necessary to define an interoperable process to virtually

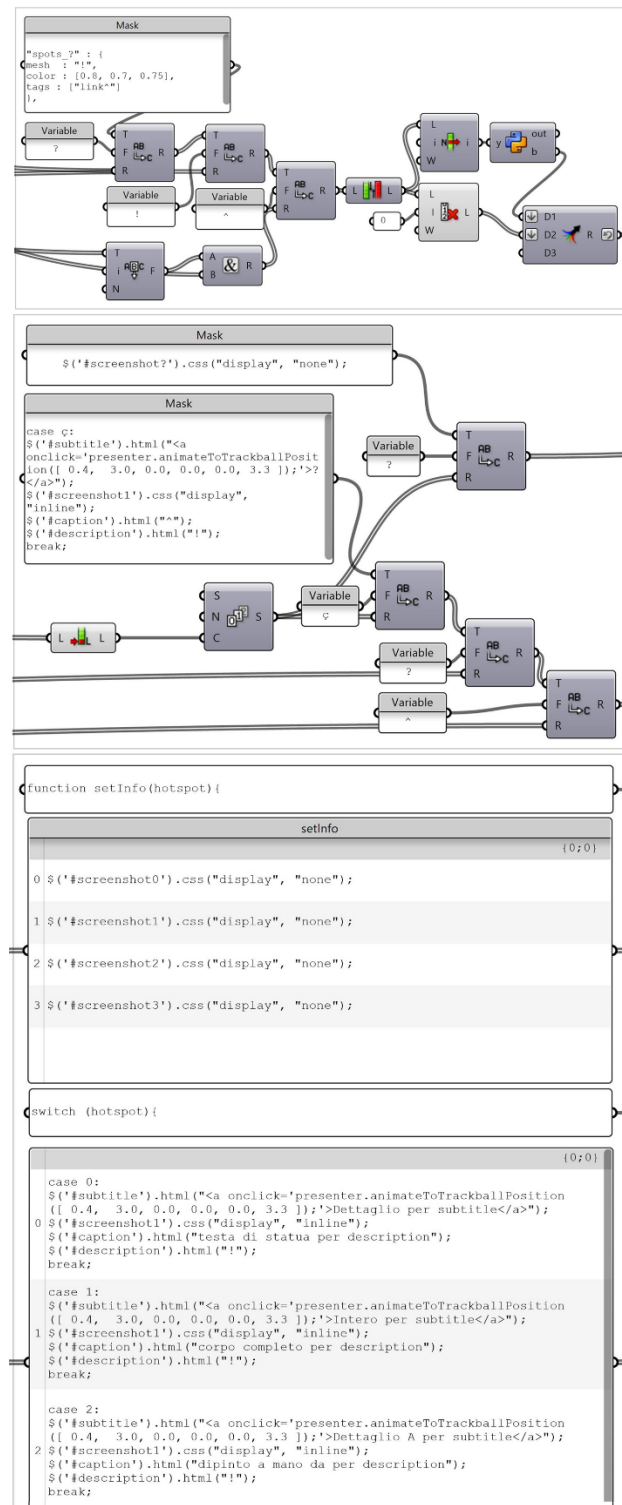


Figure 14. Textual mask structured according to the html syntax to create the consequentiality of the code parts

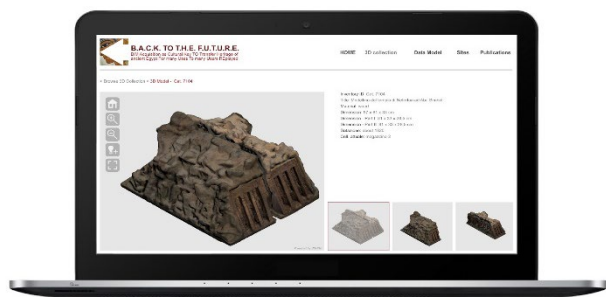


Figure 15. Interactive Web presentation of high-resolution 3D models

reproduce 3D objects by integrating geometric and semantic information, whose data can be located on different platforms. At the regard, the Visual Programming Language environment was used to import 3D models and to define the information attributes of the model or parts of it; The VPL system has been related to the virtual environment that defines its ontological structure, managed on a separate platform and subsequently published and available online through web service.

The further development of the research foresees a further extension of the project, through the development of a prototype system that virtually reproduces objects and collections (content – according to CIM process oriented) in relation to its museum (container- within the BIM environment) allowing users to operate on the system of relationships content / container of the exhibition space.

This issue is thought to support virtuous procedures of automated control of environmental requirements contained both in the object and in the building schedules, most commonly used in museums.

In this way a complex system of relationships between subjects, heritage and digital technologies is established.

ACKNOWLEDGEMENTS

The research is a part of the pilot initiative “Create a network around your research idea” is funded within the framework of the collaboration between *Politecnico di Torino* and *Compagnia di San Paolo*. The initiative is aimed at facilitating synergies among universities, companies and other entities of the socio-economic system for cooperative, multidisciplinary and innovative research projects. The research, developed in collaboration with the Museum of Egyptian Antiquities (Museo Egizio) of Turin and the University of Salamanca, aims at defining a new methodology in which the Information Modeling tools can be used for unconventional purposes, to realize 3D databases of small objects, especially those belonging to large museum collections, useful both for Scientific Research and to set up Virtual Platforms for Dissemination.

REFERENCES

- Berners-Lee, T. (2006). Linked Data - Design Issues. Retrieved December 13, 2018, from <https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, (may).
- Bruseker, G., Carboni, N., & Anais, G. (2017). *Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC*

CRM. <https://doi.org/10.1007/978-3-319-65370-9>

Coburn, E., Light, R., McKenna, G., Stein, R., & Vitzthum, A. (2010). LIDO - Lightweight Information Describing Objects Version 1.0. Retrieved from <http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html>

Di Marco, G. (2017). *Simplified complexity. Metodo per la modellazione NURBS avanzata con Rhinoceros*. Ediz. integrale. Le Pensur. Retrieved from https://books.google.it/books?id=It1_tAEACAAJ

Doerr, M., Stead, S., & Theodoridou, M. (2015). CRMdig version 3.2.1.

FORTH-ICS. (2012). 3M - Mapping Memory Manager. Retrieved from <http://139.91.183.3/3M/>

Forum on Information Standards in Heritage (FISH). (2012). MIDAS Heritage. The UK historic Environment Data Standard v1.1. Retrieved from <https://historicengland.org.uk/images-books/publications/midas-heritage/>

ICCD. (n.d.). Retrieved from <http://www.iccd.beniculturali.it/>

Le Boeuf, P., Doerr, M., Emil Ore, C., & Stead, S. (2018). CIDOC-CRM version 6.2.4. CIDOC Conceptual Reference Model (CRM).

Markoff, J. (2006). Entrepreneurs See a Web Guided by Common Sense. *The New York Times*, (11 november). Retrieved from <https://www.nytimes.com/2006/11/12/business/12web.html>

Masci, M. E. (2009). VRA Core 4.0 Descrizione degli Elementi – Traduzione italiana. Retrieved from <http://www.loc.gov/standards/vracore/schemas.html>

Meghini, C., Dekkers, M., Gradmann, S., Clayphan, R., Molendijk, J., Purday, J., & Charles, V. (2016). *Europeana Data Model Primer*. (A. Isaac, Ed.). Retrieved from <https://pro.europeana.eu/resources/standardization-tools/edm-documentation>

Migliari, R. (2008). *Geometria Descrittiva. Vol. 2*. CittàStudi.

Miles, A., & Bechhofer, S. (2009). Simple Knowledge Organization System Reference. Retrieved from <https://www.w3.org/2004/02/skos/>

O'reilly, T. (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 1(65), 17–37. Retrieved from <http://mpira.ub.uni-muenchen.de/4580/>

Powell, A., Nilsson, M., Naeve, A., Johnston, P., & Baker, T. (2007). DCMI Abstract Model. Retrieved from <http://dublincore.org/>

Tedeschi, A., & Andreani, S. (2014). *AAD, Algorithms-aided Design: Parametric Strategies Using Grasshopper*. Le Pensur Publisher. Retrieved from <https://books.google.it/books?id=VWPcoQEACAAJ>

Tillett, D. B. (2005). What is FRBR? A conceptual model for the bibliographic universe. *The Australian Library Journal*, 54(1), 24–30. <https://doi.org/10.1080/00049670.2005.10721710>

X3ML. (n.d.). Retrieved from <https://github.com/isl/x3ml>